1600

RAW SEQUENCE LISTING   DATE: 05/06/2003
PATENT APPLICATION: **US/09/870,353A**   TIME: 14:43:20

Input Set : **A:\-1-1-1.app**
Output Set: **N:\CRF4\05062003\I870353A.raw**

```
 3 <110> APPLICANT: Wang, Yan
 4       Xi, Lei
 5       Prosen, Dennis E.
 6       MJ Bioworks, Inc.
 8 <120> TITLE OF INVENTION: Improved Nucleic Acid Modifying Enzymes
10 <130> FILE REFERENCE: 020130-000111US
12 <140> CURRENT APPLICATION NUMBER: US 09/870,353A
13 <141> CURRENT FILING DATE: 2001-05-30
15 <150> PRIOR APPLICATION NUMBER: US 60/207,567
16 <151> PRIOR FILING DATE: 2000-05-26
18 <150> PRIOR APPLICATION NUMBER: US 09/640,958
19 <151> PRIOR FILING DATE: 2000-08-16
21 <160> NUMBER OF SEQ ID NOS: 34
23 <170> SOFTWARE: PatentIn Ver. 2.1
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 189
27 <212> TYPE: DNA
28 <213> ORGANISM: Artificial Sequence
30 <220> FEATURE:
31 <223> OTHER INFORMATION: Description of Artificial Sequence:synthetic
32       sequence non-specific double-stranded nucleic acid
33       binding protein Ssod7d
35 <220> FEATURE:
36 <221> NAME/KEY: CDS
37 <222> LOCATION: (1)..(189)
38 <223> OTHER INFORMATION: Ssod7d
40 <400> SEQUENCE: 1
41 gcaaccgtaa agttcaagta caaaggcgaa gaaaaagagg tagacatctc caagatcaag 60
42 aaagtatggc gtgtgggcaa gatgatctcc ttcacctacg acgagggcgg tggcaagacc 120
43 ggccgtggtg cggtaagcga aaaggacgcg ccgaaggagc tgctgcagat gctggagaag 180
44 cagaaaaag                                                      189
47 <210> SEQ ID NO: 2
48 <211> LENGTH: 63
49 <212> TYPE: PRT
50 <213> ORGANISM: Artificial Sequence
52 <220> FEATURE:
53 <223> OTHER INFORMATION: Description of Artificial Sequence:synthetic
54       sequence non-specific double-stranded nucleic acid
55       binding protein Ssod7d
57 <400> SEQUENCE: 2
58 Ala Thr Val Lys Phe Lys Tyr Lys Gly Glu Glu Lys Glu Val Asp Ile
59  1               5                  10                  15
60 Ser Lys Ile Lys Lys Val Trp Arg Val Gly Lys Met Ile Ser Phe Thr
```

ENTERED

```
61                20                   25                   30
62 Tyr Asp Glu Gly Gly Gly Lys Thr Gly Arg Gly Ala Val Ser Glu Lys
63           35                   40                   45
64 Asp Ala Pro Lys Glu Leu Leu Gln Met Leu Glu Lys Gln Lys Lys
65      50                   55                   60
68 <210> SEQ ID NO: 3
69 <211> LENGTH: 1899
70 <212> TYPE: DNA
71 <213> ORGANISM: Artificial Sequence
73 <220> FEATURE:
74 <223> OTHER INFORMATION: Description of Artificial Sequence:fusion protein
75      Ssod7d-deltaTaq
77 <220> FEATURE:
78 <221> NAME/KEY: CDS
79 <222> LOCATION: (1)..(1899)
80 <223> OTHER INFORMATION: Ssod7d-deltaTaq
82 <400> SEQUENCE: 3
83 atgattacga attcgagcgc aaccgtaaag ttcaagtaca aaggcgaaga aaaagaggta 60
84 gacatctcca agatcaagaa agtatggcgt gtgggcaaga tgatctcctt cacctacgac 120
85 gagggcggtg gcaagaccgg ccgtggtgcg gtaagcgaaa aggacgcgcc gaaggagctg 180
86 ctgcagatgc tggagaagca gaaaaagggc ggcggtgtca ctagtcccaa ggccctggag 240
87 gaggcccct ggcccccgcc ggaaggggcc ttcgtgggct ttgtgctttc ccgcaaggag 300
88 cccatgtggg ccgatcttct ggccctggcc gccgccaggg ggggccgggt ccaccgggcc 360
89 cccgagcctt ataaagccct cagggacctg aaggaggcgc gggggcttct cgccaaagac 420
90 ctgagcgttc tggccctgag ggaaggcctt ggcctcccgc ccggcgacga ccccatgctc 480
91 ctcgcctacc tcctggaccc ttccaacacc accccccgagg gggtggcccg gcgctacggc 540
92 ggggagtgga cggaggaggc gggggagcgg gccgcccttt ccgagaggct cttcgccaac 600
93 ctgtggggga ggcttgaggg ggaggagagg ctcctttggc tttaccggga ggtggagagg 660
94 ccccttttccg ctgtcctggc ccacatggag gccacggggg tgcgcctgga cgtggcctat 720
95 ctcagggcct tgtccctgga ggtggccgag gagatcgccc gcctcgaggc cgaggtcttc 780
96 cgcctggccg gccacccctt caacctcaac tcccgggacc agctggaaag ggtcctcttt 840
97 gacgagctag ggcttcccgc catcggcaag acggagaaga ccggcaagcg ctccaccagc 900
98 gccgccgtcc tggaggccct ccgcgaggcc caccccatcg tggagaagat cctgcagtac 960
99 cgggagctca ccaagctgaa gagcacctac attgacccct gccggacct catccacccc 1020
100 aggacgggcc gcctccacac ccgcttcaac cagacggcca cggccacggg caggctaagt 1080
101 agctccgatc ccaacctcca gaacatcccc gtccgcaccc cgcttgggca gaggatccgc 1140
102 cgggccttca tcgccgagga ggggtggcta ttggtggccc tggactatag ccagatagag 1200
103 ctcagggtgc tggcccacct ctccggcgac gagaacctga tccgggtctt ccaggagggg 1260
104 cgggacatcc acacggagac cgccagctgg atgttcggcg tccccccggga ggccgtggac 1320
105 cccctgatgc gccgggcggc caagaccatc aacttcgggg tcctctacgg catgtcggcc 1380
106 caccgcctct cccaggagct agccatccct tacgaggagg cccaggcctt cattgagcgc 1440
107 tactttcaga gcttccccaa ggtgcgggcc tggattgaga agaccctgga ggagggcagg 1500
108 aggcggggggt acgtggagac cctcttcggc cgccgccgct acgtgccaga cctagaggcc 1560
109 cgggtgaaga gcgtgcggga ggcggccgag cgcatggcct tcaacatgcc cgtccagggc 1620
110 accgccgccg acctcatgaa gctggctatg gtgaagctct tccccaggct ggaggaaatg 1680
111 ggggccagga tgctccttca ggtccacgac gagctggtcc tcgaggcccc aaaagagagg 1740
112 gcggaggccg tggcccggct ggccaaggag gtcatggagg gggtgtatcc cctggccgtg 1800
113 cccctggagg tggaggtggg gataggggag gactggctct ccgccaagga gggcattgat 1860
114 ggccgcggcg gaggcgggca tcatcatcat catcattaa 1899
```

```
117 <210> SEQ ID NO: 4
118 <211> LENGTH: 632
119 <212> TYPE: PRT
120 <213> ORGANISM: Artificial Sequence
122 <220> FEATURE:
123 <223> OTHER INFORMATION: Description of Artificial Sequence:fusion protein
124      Ssod7d-deltaTaq
126 <400> SEQUENCE: 4
127 Met Ile Thr Asn Ser Ser Ala Thr Val Lys Phe Lys Tyr Lys Gly Glu
128   1               5                  10                  15
129 Glu Lys Glu Val Asp Ile Ser Lys Ile Lys Lys Val Trp Arg Val Gly
130              20                  25                  30
131 Lys Met Ile Ser Phe Thr Tyr Asp Glu Gly Gly Gly Lys Thr Gly Arg
132          35                  40                  45
133 Gly Ala Val Ser Glu Lys Asp Ala Pro Lys Glu Leu Leu Gln Met Leu
134      50                  55                  60
135 Glu Lys Gln Lys Lys Gly Gly Gly Val Thr Ser Pro Lys Ala Leu Glu
136  65                  70                  75                  80
137 Glu Ala Pro Trp Pro Pro Pro Glu Gly Ala Phe Val Gly Phe Val Leu
138              85                  90                  95
139 Ser Arg Lys Glu Pro Met Trp Ala Asp Leu Leu Ala Leu Ala Ala Ala
140              100                 105                 110
141 Arg Gly Gly Arg Val His Arg Ala Pro Glu Pro Tyr Lys Ala Leu Arg
142          115                 120                 125
143 Asp Leu Lys Glu Ala Arg Gly Leu Leu Ala Lys Asp Leu Ser Val Leu
144      130                 135                 140
145 Ala Leu Arg Glu Gly Leu Gly Leu Pro Pro Gly Asp Asp Pro Met Leu
146 145                 150                 155                 160
147 Leu Ala Tyr Leu Leu Asp Pro Ser Asn Thr Thr Pro Glu Gly Val Ala
148              165                 170                 175
149 Arg Arg Tyr Gly Gly Glu Trp Thr Glu Glu Ala Gly Glu Arg Ala Ala
150          180                 185                 190
151 Leu Ser Glu Arg Leu Phe Ala Asn Leu Trp Gly Arg Leu Glu Gly Glu
152          195                 200                 205
153 Glu Arg Leu Leu Trp Leu Tyr Arg Glu Val Glu Arg Pro Leu Ser Ala
154      210                 215                 220
155 Val Leu Ala His Met Glu Ala Thr Gly Val Arg Leu Asp Val Ala Tyr
156 225                 230                 235                 240
157 Leu Arg Ala Leu Ser Leu Glu Val Ala Glu Glu Ile Ala Arg Leu Glu
158              245                 250                 255
159 Ala Glu Val Phe Arg Leu Ala Gly His Pro Phe Asn Leu Asn Ser Arg
160              260                 265                 270
161 Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu Gly Leu Pro Ala Ile
162          275                 280                 285
163 Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser Thr Ser Ala Ala Val Leu
164      290                 295                 300
165 Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln Tyr
166 305                 310                 315                 320
167 Arg Glu Leu Thr Lys Leu Lys Ser Thr Tyr Ile Asp Pro Leu Pro Asp
```

RAW SEQUENCE LISTING   DATE: 05/06/2003
·PATENT APPLICATION: **US/09/870,353A**   TIME: 14:43:20

Input Set : **A:\-1-1-1.app**
Output Set: **N:\CRF4\05062003\I870353A.raw**

```
168                 325                 330                 335
169 Leu Ile His Pro Arg Thr Gly Arg Leu His Thr Arg Phe Asn Gln Thr
170             340                 345                 350
171 Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp Pro Asn Leu Gln Asn
172         355                 360                 365
173 Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe Ile
174     370                 375                 380
175 Ala Glu Glu Gly Trp Leu Leu Val Ala Leu Asp Tyr Ser Gln Ile Glu
176 385             390                 395                 400
177 Leu Arg Val Leu Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg Val
178             405                 410                 415
179 Phe Gln Glu Gly Arg Asp Ile His Thr Glu Thr Ala Ser Trp Met Phe
180             420                 425                 430
181 Gly Val Pro Arg Glu Ala Val Asp Pro Leu Met Arg Arg Ala Ala Lys
182         435                 440                 445
183 Thr Ile Asn Phe Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu Ser
184     450                 455                 460
185 Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Gln Ala Phe Ile Glu Arg
186 465                 470                 475                 480
187 Tyr Phe Gln Ser Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr Leu
188             485                 490                 495
189 Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu Thr Leu Phe Gly Arg Arg
190         500                 505                 510
191 Arg Tyr Val Pro Asp Leu Glu Ala Arg Val Lys Ser Val Arg Glu Ala
192         515                 520                 525
193 Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln Gly Thr Ala Ala Asp
194     530                 535                 540
195 Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro Arg Leu Glu Glu Met
196 545                 550                 555                 560
197 Gly Ala Arg Met Leu Leu Gln Val His Asp Glu Leu Val Leu Glu Ala
198             565                 570                 575
199 Pro Lys Glu Arg Ala Glu Ala Val Ala Arg Leu Ala Lys Glu Val Met
200         580                 585                 590
201 Glu Gly Val Tyr Pro Leu Ala Val Pro Leu Glu Val Glu Val Gly Ile
202     595                 600                 605
203 Gly Glu Asp Trp Leu Ser Ala Lys Glu Gly Ile Asp Gly Arg Gly Gly
204     610                 615                 620
205 Gly Gly His His His His His His
206 625                 630
```

209 <210> SEQ ID NO: 5
210 <211> LENGTH: 2763
211 <212> TYPE: DNA
212 <213> ORGANISM: Artificial Sequence
214 <220> FEATURE:
215 <223> OTHER INFORMATION: Description of Artificial Sequence:fusion protein
216   Ssod7d/full-length Taq
218 <220> FEATURE:
219 <221> NAME/KEY: CDS
220 <222> LOCATION: (1)..(2763)

```
221 <223> OTHER INFORMATION: Ssod7d/full-length Taq
223 <400> SEQUENCE: 5
224 atgattacga attcgagcgc aaccgtaaag ttcaagtaca aaggcgaaga aaaagaggta 60
225 gacatctcca agatcaagaa agtatggcgt gtgggcaaga tgatctcctt cacctacgac 120
226 gagggcggtg gcaagaccgg ccgtggtgcg gtaagcgaaa aggacgcgcc gaaggagctg 180
227 ctgcagatgc tggagaagca gaaaaagggc ggcggtgtca ctagtgggat gctgcccctc 240
228 tttgagccca agggccgggt cctcctggtg gacggccacc acctggccta ccgcaccttc 300
229 cacgccctga agggcctcac caccagccgg ggggagccgg tgcaggcggt ctacggcttc 360
230 gccaagagcc tcctcaaggc cctcaaggag gacggggacg cggtgatcgt ggtctttgac 420
231 gccaaggccc cctccttccg ccacgaggcc tacggggggt acaaggcggg ccgggccccc 480
232 acgccagagg actttccccg gcaactcgcc ctcatcaagg agctggtgga cctcctgggg 540
233 ctggcgcgcc tcgaggtccc gggctacgag gcggacgacg tcctggccag cctggccaag 600
234 aaggcggaaa aggagggcta cgaggtccgc atcctcaccg ccgacaaaga cctttaccag 660
235 ctcctttccg accgcatcca cgtcctccac cccgagggg t acctcatcac cccggcctgg 720
236 ctttgggaaa agtacggcct gaggcccgac cagtgggccg actaccgggc cctgaccggg 780
237 gacgagtccg acaaccttcc cggggtcaag ggcatcgggg agaagacggc gaggaagctt 840
238 ctggaggagt ggggggagcct ggaagccctc ctcaagaacc tggaccggct gaagcccgcc 900
239 atccgggaga agatcctggc ccacatggac gatctgaagc tctcctggga cctggccaag 960
240 gtgcgcaccg acctgcccct ggaggtggac ttcgccaaaa ggcgggagcc cgaccgggag 1020
241 aggcttaggg cctttctgga gaggcttgag tttggcagcc tcctccacga gttcggcctt 1080
242 ctggaaagcc ccaaggccct ggaggaggcc ccctggcccc cgccggaagg ggccttcgtg 1140
243 ggctttgtgc tttcccgcaa ggagcccatg tgggccgatc ttctggcccct ggccgccgcc 1200
244 aggggggggcc gggtccaccg gcccccccgag ccttataaag ccctcaggga cctgaaggag 1260
245 gcgcggggggc ttctcgccaa agacctgagc gttctggccc tgagggaagg ccttggcctc 1320
246 ccgcccggcg acgacccccat gctcctcgcc tacctcctgg acccttccaa caccacccc c 1380
247 gaggggggtgg cccggcgcta cggcggggag tggacggagg aggcgggggga gcgggccgcc 1440
248 ctttccgaga ggctcttcgc caacctgtgg ggggaggcttg aggggggagga gaggctcctt 1500
249 tggctttacc gggaggtgga gaggcccctt tccgctgtcc tggcccacat ggaggccacg 1560
250 ggggtgcgcc tggacgtggc ctatctcagg gccttgtccc tggaggtggc cgaggagatc 1620
251 gccccgcctcg aggccgaggt cttccgcctg gccggccacc ccttcaacct caactcccgg 1680
252 gaccagctgg aaagggtcct cttttgacgag ctagggcttc ccgccatcgg caagacggag 1740
253 aagaccggca agcgctccac cagcgccgcc gtcctggagg ccctccgcga ggcccacccc 1800
254 atcgtggaga agatcctgca gtaccgggag ctcaccaagc tgaagagcac ctacattgac 1860
255 cccttgccgg acctcatcca ccccaggacg ggccgcctcc acacccgctt caaccagacg 1920
256 gccacggcca cgggcaggct aagtagctcc gatcccaacc tccagaacat ccccgtccgc 1980
257 accccgcttg ggcagaggat ccgccgggcc ttcatcgccg aggaggggtg gctattggtg 2040
258 gccctggact atagccagat agagctcagg gtgctggccc acctctccgg cgacgagaac 2100
259 ctgatccggg tcttccagga ggggcgggac atccacacgg agaccgccag ctggatgttc 2160
260 ggcgtccccc gggaggccgt ggacccccctg atgcgccggg cggccaagac catcaacttc 2220
261 ggggtcctct acggcatgtc ggcccaccgc ctctcccagg agctagccat cccttacgag 2280
262 gaggcccagg ccttcattga gcgctacttt cagagcttcc ccaaggtgcg ggcctggatt 2340
263 gagaagaccc tggaggaggg caggaggcgg gggtacgtgg agaccctctt cggccgccgc 2400
264 cgctacgtgc cagacctaga ggcccggggtg aagagcgtgc gggaggcggc cgagcgcatg 2460
265 gccttcaaca tgcccgtcca gggcaccgcc gccgacctca tgaagctggc tatggtgaag 2520
266 ctcttcccca ggctggagga aatggggggcc aggatgctcc ttcaggtcca cgacgagctg 2580
267 gtcctcgagg ccccaaaaga gagggcggag ccgtggcccc ggctggccaa ggaggtcatg 2640
268 gagggggtgt atcccctggc cgtgcccctg gaggtggagg tggggatagg ggaggactgg 2700
269 ctctccgcca aggagggcat tgatggccgc ggcggaggcg ggcatcatca tcatcatcat 2760
270 taa                                                          2763
```

VERIFICATION SUMMARY   DATE: 05/06/2003
PATENT APPLICATION: **US/09/870,353A**   TIME: 14:43:21

Input Set : **A:\-1-1-1.app**
Output Set: **N:\CRF4\05062003\I870353A.raw**